Complete and assembled genome sequence of Salmonella enterica subsp. enterica serotype Senftenberg N17-509, a strain lacking Salmonella pathogen island SPI-1

Stevens, Marc J A ; Zurfluh, Katrin ; Althaus, D ; Corti, S ; Lehner, Angelika ; Stephan, Roger

Abstract: The genome of Salmonella enterica subsp. enterica serotype Senftenberg N17-509, a strain isolated from desiccated coconut, was sequenced using single-molecule real-time sequencing. It consists of a 5.1-Mbp chromosome and a 29-kb linear plasmid.

DOI: https://doi.org/10.1128/genomeA.00156-18

Complete and assembled genome sequence of *Salmonella enterica* subsp. *enterica* serotype Senftenberg N17-509, a strain lacking *Salmonella* pathogen island SPI-1

Marc J. A. Stevens, Katrin Zurfluh, Denise Althaus, Sabrina Corti, Angelika Lehner, Roger Stephan*

Institute for Food Safety and Hygiene, Vetsuisse Faculty, University of Zurich, Switzerland.

*Corresponding author: roger.stephan@uhz.ch

Abstract.

The genome of *Salmonella enterica* subsp. *enterica* serotype Senftenberg N17-509, a strain isolated from desiccated coconut, was sequenced using single molecule real time sequencing. It consists of a 5.1-Mbp chromosome, and a 29-kb linear plasmid.

The *Salmonella enterica* subsp. *enterica* strain N17-509 was isolated from desiccated coconut and provided different identification and typing challenges. The biochemical tests performed to confirm the *Salmonella* genus were inconclusive. A commercial real-time PCR identification system (Assurance GDS® *Salmonella* kit (BioControl Systems, Bellevue, WA, USA)) based on the *invA* gene yielded a negative result. A further rRNA based fast method (VIT Salmonella kit (vermicon, München, Germany) as well as the

MALDI TOF based identification was positive for *Salmonella*. According to the KAUFFMANN-WHITE Scheme (Institut Pasteur, Paris), the strain was serotyped as *Salmonella enterica* subsp. *enterica* serovar 1,3,19:nt.

Genomic DNA was extracted using Wizard® Genomic DNAPurification Kit according to the manufacturer's protocol (Promega AG, Dübendorf, Switzerland). The genome was sequenced using a two SMRT cells on a PacBio RS II (Pacific Biosciences, Menlo Park, CA, USA) and was performed at the Functional Genomics Center Zurich (Zurich, Switzerland). The raw reads were filtered using the RS_Filter_Only protocol in the SMRT-portal (Pacific Biosciences) with 0.75 as polymerase read quality cut off and a minimal length of 500 bp. A total of 91,709 reads with an average length of 11,001 bp were selected, corresponding to 1,008,898,178 sequenced base pairs and a 197-fold coverage. The reads were assembled using the Canu assembler 1.6 (1) with the option "pacbio-raw" and an estimated genome size of 5.1 Mbp. The Canu output consisted of 3 contigs, which were further polished in CLC workbench 7 (CLC, Aarhus, Denmark). Two contigs were combined and formed the chromosomal sequence. The third contig contained a plasmid-like sequence. The start of the chromosome was set at the origin of replication, determined using OriCfinder (2).

A blastN with the complete chromosome against all *Enterobacteriales* genomes in the NCBI nucleotide database revealed highest identity of 99% and a coverage of 91% with the genome of *Salmonella enterica* subsp. *enterica* serotype Senftenberg 775W (=ATCC 43845). Hence, strain N17-509 was classified as *Salmonella enterica* subsp. *enterica* serotype Senftenberg. The genome of N17-509 consists of a 5,121,989-bp chromosome and a 29,600-bp extra-chromosomal element designated pN17-509. The GC-contents of

the chromosome and pN17-509 are 52.2% and 36,5%, respectively. The genome was annotated by the NCBI Prokaryotic Genomes Automatic Annotation Pipeline and contains 5361 genes of which 5241 encoding proteins. Further, 84 tRNA genes and 7 rRNA operons are present. Remarkably, the *Salmonella* pathogen island SPI-1, harbouring e.g. the *invA* gene, is absent in this strain. The plasmid pN17-509 has 97% identity at 87% coverage with the linear plasmid pBSSB1 from *Salmonella enterica* subsp. *enterica* serovar Typhi that mediates flagellar variation (3).

Nucleotide sequence accession number. Sequence and annotation data of the complete genome of *Salmonella enterica* subsp. *enterica* serotype Senftenberg N17-509 are deposited in the GenBank database with the accession number CP026379 for the chromosome and CP026380 for the plasmid.

REFERENCES

1) **Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM.** 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res **27**:722-736.

2) **Gao F, Zhang CT.** 2008. Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. BMC Bioinformatics **9**:79.

3) **Baker S, Hardy J, Sanderson KE, Quail M, Goodhead I, Kingsley RA, Parkhill J, Stocker B, Dougan G.** 2007. A novel linear plasmid mediates flagellar variation in *Salmonella* Typhi. PLoS Pathog **11**;3(5):e59.